Administrative Records Experiment in 2000 (AREX 2000) Outcomes Evaluation

APPENDICES

APPENDIX 1: PROFILE OF TEST SITES

A1.1 County demographics

The ability to accurately measure the resident population using administrative records is likely to vary by the age, race, sex, and Hispanic composition of the AREX counties. These demographic groups are likely to have distinct coverage rates within administrative records, as well as mobility, fertility, and mortality rates. The latter rates are also likely to interact with the record-keeping processes of the federal agencies that collect and maintain the data. The sites were chosen for their varying demographic characteristics to test the feasibility of enumerating the population using administrative records. Table A1.1 provides a detailed breakdown of 2000 demographic characteristics for the five counties in the AREX test sites. Some general comments on the AREX test sites include:

- Baltimore and Baltimore City have the largest populations, compared to the less populated CO counties.
- Females exceed males in all five counties; the sex ratio is larger in the CO counties.
- The MD counties are much older than the CO counties; the age 0-4 age group proportions are larger in the CO counties, while the older age groups are larger in the MD counties.
- Baltimore City, and to a lesser extent, Baltimore County, have large Black populations; Hispanics are the largest minority population in CO, followed by APIs.

Table A1: Demographic Breakdown of the Census 2000 Household Population for AREX Counties

	Baltimore County		Baltimore City		Douglas County		El Paso County		Jefferson County	
Total	736,652		625,401		175,300		501,533		519,326	
White	548,776	74.5%	196,427	31.4%	162,639	92.8%	408,167	81.4%	471,107	90.7%
Black	147,226	20.0%	404,198	64.6%	1,663	0.9%	31,875	6.4%	4,126	0.8%
AI	1,923	0.3%	2,097	0.3%	716	0.4%	4,725	0.9%	3,971	0.8%
API	23,631	3.2%	9,168	1.5%	4,488	2.6%	13,954	2.8%	12,330	2.4%
Hispanic	13,433	1.8%	10,712	1.7%	8,825	5.0%	56,677	11.3%	51,346	9.9%
Age 0-4	45,179	6.1%	41,593	6.7%	16,949	9.7%	39,006	7.8%	33,213	6.4%
5-19	147,393	20.0%	135,558	21.7%	41,376	23.6%	115,404	23.0%	111,655	21.5%
20-24	41,740	5.7%	43,627	7.0%	5,478	3.1%	32,596	6.5%	28,901	5.6%
25-34	100,363	13.6%	89,525	14.3%	28,552	16.3%	75,205	15.0%	70,672	13.6%
35-44	122,116	16.6%	97,983	15.7%	38,007	21.7%	90,039	18.0%	96,357	18.6%
45-54	107,499	14.6%	81,691	13.1%	26,235	15.0%	68,878	13.7%	84,174	16.2%
55-64	67,187	9.1%	53,630	8.6%	11,597	6.6%	37,709	7.5%	46,190	8.9%
65+	105,175	14.3%	81,794	13.1%	7,106	4.1%	42,696	8.5%	48,164	9.3%
65-74	54,768	7.4%	43,533	7.0%	4,784	2.7%	24,988	5.0%	28,025	5.4%
75-84	40,114	5.4%	29,618	4.7%	1,959	1.1%	14,211	2.8%	15,900	3.1%
85+	10,293	1.4%	8,643	1.4%	363	0.2%	3,497	0.7%	4,239	0.8%
Male	349,319	47.4%	288,070	46.1%	87,478	49.9%	248,764	49.6%	257,876	49.7%
Female	387,333	52.6%	337,331	53.9%	87,822	50.1%	252,769	50.4%	261,450	50.3%

A1.2 Spatial and ecological issues affecting AREX tracts

Summary: Though it appears that tracts with moderate/high population density have more vacant and/or rental units, this is not true for all tracts in the MD and CO AREX counties. Some higher density tracts may have more desirable neighborhoods and fewer vacant units. Similarly, there is evidence that suburban and rural tracts may have less stable net migration of residents. In some cases, new home construction may be related to vacant units, however, the spatial maps do not identify new home subdivisions.

Figure A1.1a: Number of Vacant Housing Units: MD Tracts

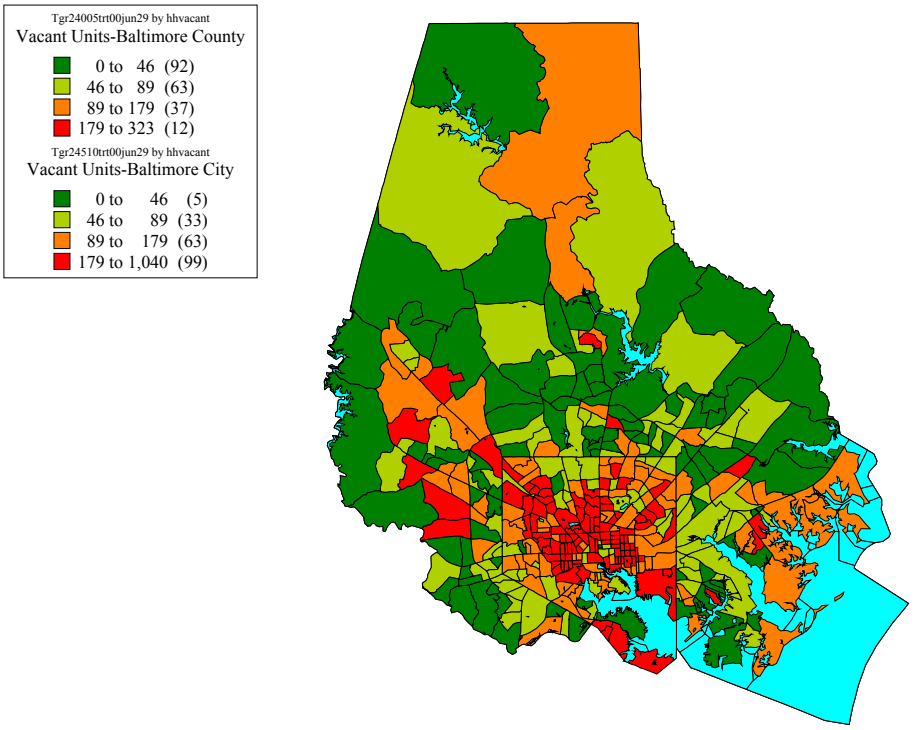


Figure A1.1b: Number of Vacant Housing Units: CO Tracts

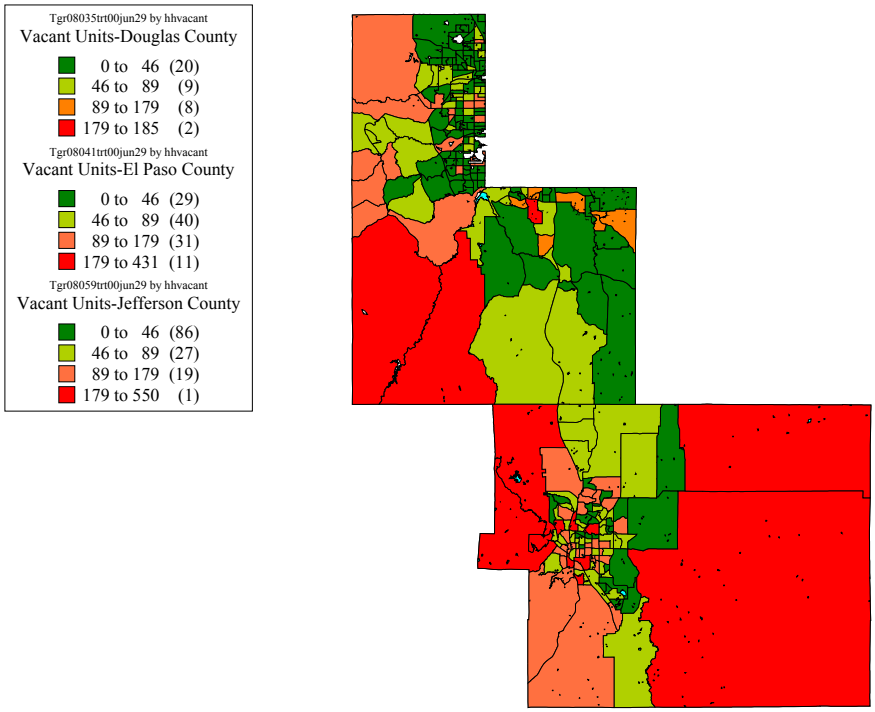


Figure A1.2a: Population Density: MD Tracts

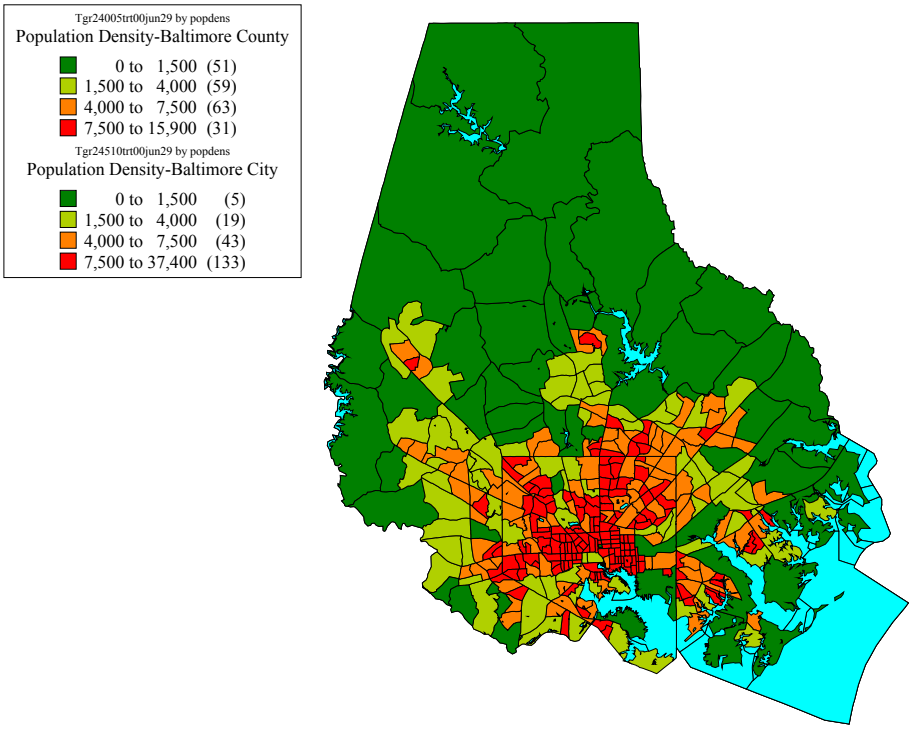
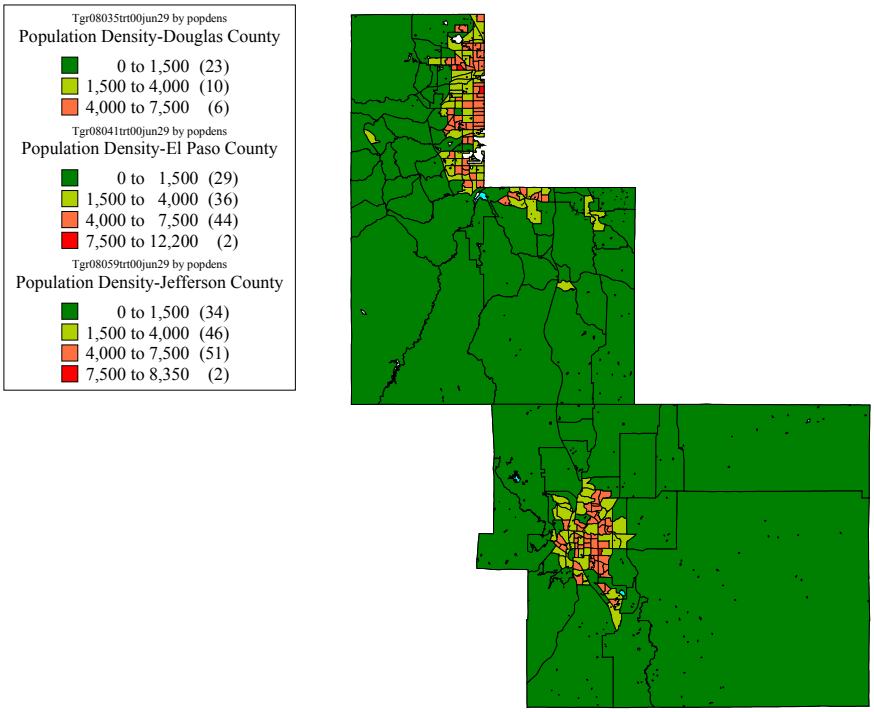


Figure A1.2b: Population Density: CO Tracts



Figures A1.1a-b, A1.2a-b show the ecological distribution of vacant housing units and population density of MD and CO tracts from Census 2000 results. These basic ecological maps suggest that tracts with larger proportions of vacant and/or rental (high-density) units are heterogeneously distributed across the AREX counties. Some of the highlights include:

- Tracts with large numbers of vacant units coincide with high-density population tracts, though this is not true for all tracts, especially around Denver in Jefferson County.
- A large number of tracts have vacant housing units, especially in downtown Baltimore City, with several tracts having clusters of moderate and high numbers of vacancies in Baltimore County.
- Despite the large land area of the CO tracts, there are few tracts with large numbers of vacant housing units; most of the vacant units are in El Paso county, within and around Colorado Springs.

Spatial and ecological issues impact how well administrative records accurately measure the resident populations of sub-county regions and their proximity to each other, and can have a variable affect on demographic group counts. Counties with a large number of vacant housing units are likely to provide poorer estimates because of the reporting lag between a moving household and federal agencies recording of population mobility. Residents of these areas may be less affluent and potentially less-covered populations. Similarly, transient population groups, like college students and military personnel, can flow into and out of other residences and group quarters. Older residents, and especially women, are more likely to enter or exit nursing homes, compared to the general population. This group also experiences higher mortality rates that may impact their coverage, due to reporting lag in recording mobility or deaths.

A1.3 Demographic diversity of AREX tracts

Summary: Age diversity is greater in urban and suburban tracts of MD, while race/Hispanic diversity is greater in urban and suburban tracts of CO. The Black population in Baltimore City is highly segregated and appears to be as homogeneous as mostly White tracts in the other counties. Some tract counts are harder to measure accurately, particularly those where multi-race reporting occurs and large numbers of non-relative household members live (not shown). These harder to measure attributes tend to affect the same tracts.

Figure A1.3a: Shannon-Wiener Diversity Index for Age-MD Tracts

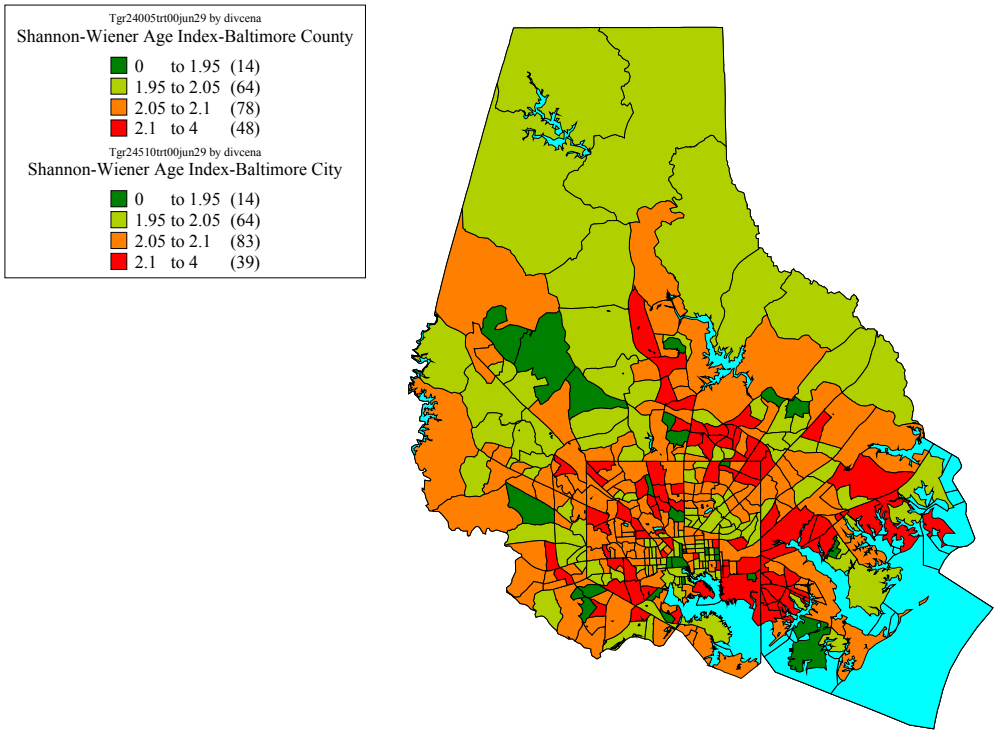


Figure A1.3b: Shannon-Wiener Diversity Index for Age-CO Tracts

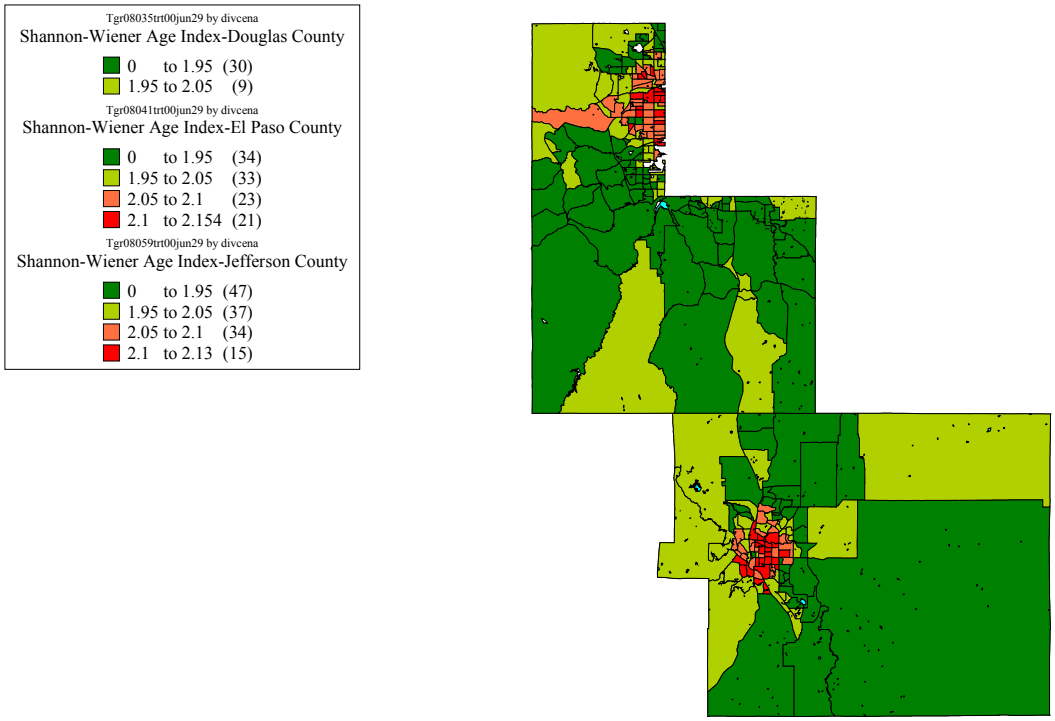


Figure A1.4a: Shannon-Wiener Diversity Index for Race-MD Tracts

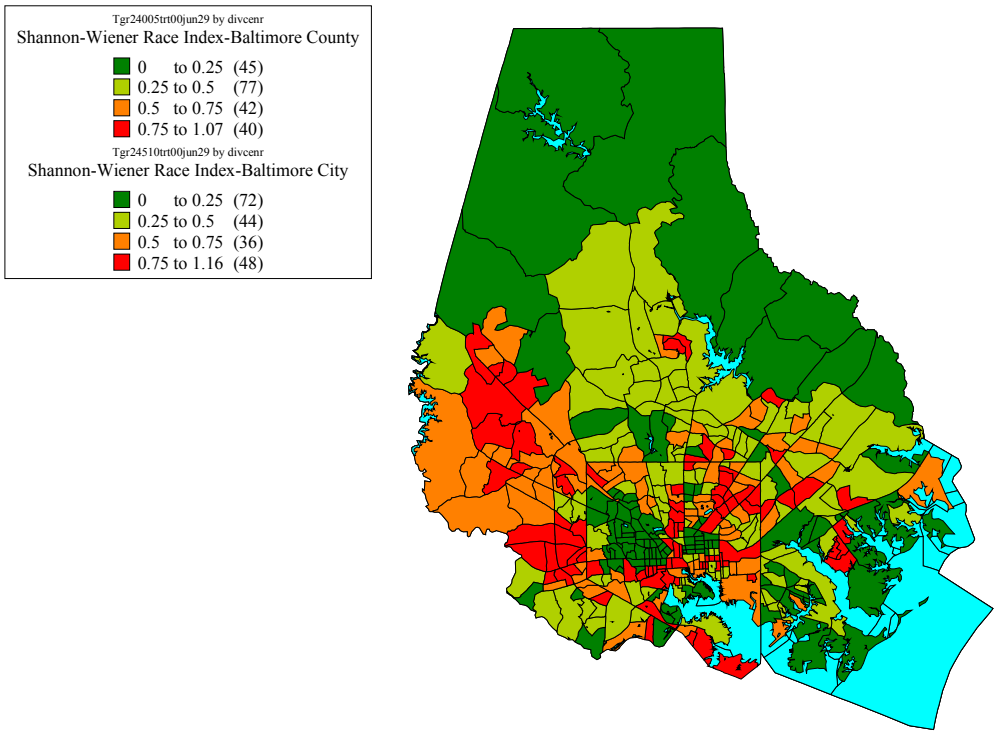
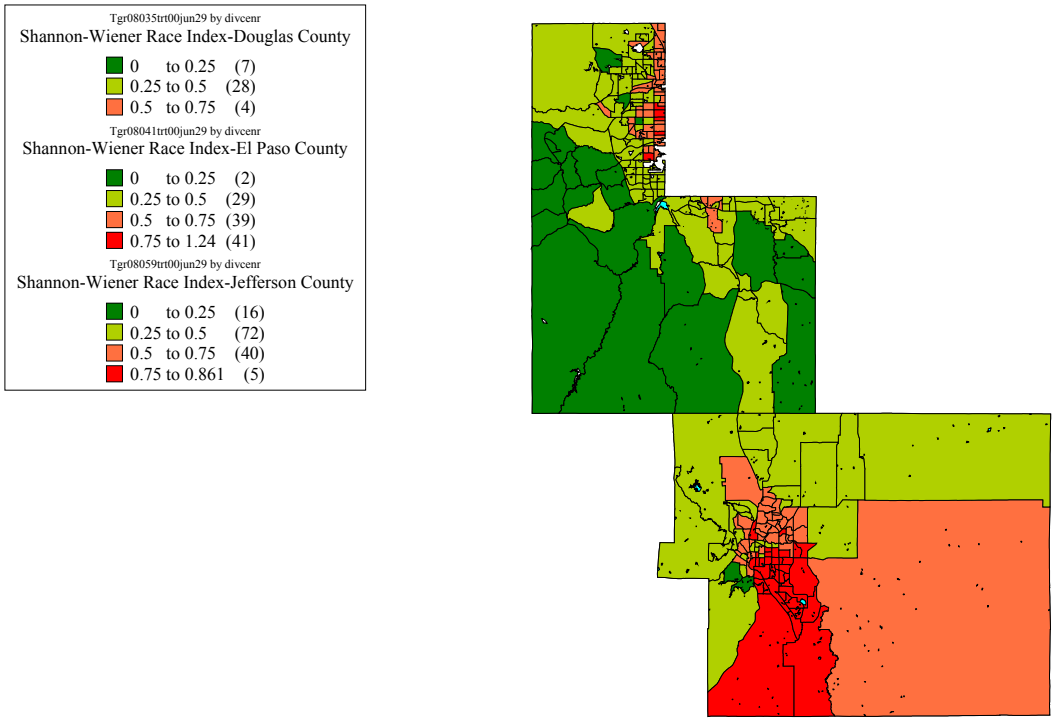


Figure A1.4b: Shannon-Wiener Diversity Index for Race-CO Tracts



The demographic characteristics of tract residents and the type, price, and availability of housing units are likely to attract or repel new in-migrants and affect tract-level coverage rates. The Shannon-Wiener diversity index measures the number of race/Hispanic groups and their population proportions within a tract, but does not distinguish whether a tract is predominantly White or Black. Tract-level diversity using Census 2000 results is shown above in Figures A1.3a-b, A1.4a-b and can be summarized:

- In the MD counties, the most diverse tracts exist in the southern, more urban section of MD; the western portion of Baltimore City with a large proportion of Blacks appears to be as racially uniform as the White, northern portion of Baltimore County.
- In the CO counties, diversity is concentrated in urban areas and several bordering tracts; this pattern may also reflect tracts with a large proportion of Hispanics and smaller White minority.

APPENDIX 2: RACE IMPUTATION

General description of the race imputation process

The race imputation process used logistic model results estimated from linked CPS-SSA Numident files, as well as Hispanic and Asian surname files and IHS records (see Bye, 1998 for complete details). The general model algorithm used the Numident, IHS, and surname identifiers to predict the matched CPS race codes. The type of Numident record, frequency of consistent race reports, geographic identifiers, and foreign birth indicators were also for calculating race probabilities. The calculated probabilities were then processed through a hot deck procedure for the final race assignment.

Persons under the age of 18 frequently lacked complete information and had blank race assignments in their Numident records. More problematic is that CPS did not include persons under age 15 years and the original model results did not address this younger age group. Consequently, the race information was incomplete and potentially inaccurate for minor children and a second stage imputation process was applied. The derived race assignment of the primary tax filer was applied to all children. While this second stage may address problems with children's records, it may also assign race from inaccurate race identifiers of some householders.

Table A2 provides the results of the race assignment process and imputed race codes by type of assignment:

Table A2: Race Assignment and Imputation Rates by Method, Race, and County

Imputation Method Most Frequent Report[1]	Baltimore County	Baltimore	Douglas	El Paso	Jefferson
All Persons	81.0%	74.8%	69.0%	72.9%	75.8%
White	82.5%	75.2%	40.2%	74.6%	77.6%
Black	79.9%	75.8%	61.3%	77.2%	59.5%
AI	55.7%	54.1%	35.5%	38.3%	36.3%
API	64.4%	56.4%	56.7%	64.5%	61.2%
Hispanic	1.2%	2.1%	1.8%	5.5%	4.4%
Imputed Primary Tax Filer Race (applied to persons under 18)[2]					
All Persons	9.4%	7.9%	13.1%	10.1%	9.7%
White	9.1%	6.7%	13.3%	10.2%	9.8%
Black	10.8%	8.6%	13.2%	12.0%	10.1%
AI	8.3%	7.5%	11.2%	9.9%	9.0%
API	8.8%	4.9%	10.3%	9.2%	9.7%
Hispanic	-	-	-	-	-

PCF Probability Model (applied to all adults)[2]

All Persons	3.1%	1.8%	4.1%	6.8%	6.3%
White	2.9%	4.3%	3.6%	6.7%	5.7%
Black	0.9%	0.3%	13.6%	3.3%	21.5%
AI	20.5%	16.7%	15.1%	10.3%	12.1%
API	19.9%	15.8%	17.9%	18.1%	21.5%
Hispanic	92.5%	82.6%	84.6%	85.3%	88.2%

[1]Most frequent race report / total AREX records
[2]Imputed records / total AREX records

APPENDIX 3: TRACT AND BLOCK INCONGRUITIES

Technical factors affecting tract and block differences

The relationship between level of geography and the accuracy of AREX counts is more complicated than it appears. For total population counts, county-level results can be hypothesized as more accurate than tract-level results, which in turn are expected to be more accurate than block-level results. And this relationship was supported by total population values across the geographic levels. However, statistical, computational, and substantive issues affect this relationship when looking at sparse populations that are likely to be distributed in a heterogeneous fashion across counties.

Table A3.1 (next page) is a listing of blocks for a single tract that focuses on AI residents and indicated AREX overcounted Census by 250 percent.[1] Each record shows the block level Algebraic Percent Error (ALPE) and AREX and census counts and difference for that block. This single tract covers 34 blocks, but only three have AI residents, based on Census results, while AREX indicates one block has AI residents. However, there are four blocks with AI residents, according to AREX, but three are zero-blocks for Census. Because of the computational problems, the block level results have two blocks each with 100 percent undercounts of census. But the five AREX persons who were not counted at the block-level contributed to a 267 percent overcount at the tract-level (11-3)/3.[2]

There is reason to be skeptical about the validity of the AREX overcounts for Census zero blocks. AREX overcounts may indicate a single person in a block is an AI but one would expect at least two or three AIs in a block, reflecting family members and neighbors with similar backgrounds living in the same neighborhood. The validity of these overcounts is important when considering the accuracy of the various geographic levels. One would expect the greatest accuracy at the county-level, because AREX overcounts could be 'absorbed' by the larger population counts. At the tract level, AREX overcounts are included in calculations, but tract-level denominators are sometimes small, resulting in inflated ALPE overcounts and highly skewed distributions that are sometimes U-shaped. At the block-level, AREX overcounts are not included in the distributions and calculations because the zero-blocks render these as undefined. This is problematic for small populations and sparse distributions, especially AIs and persons 75+ or 85+.

[1] Actual tract numbers have been dummied to ensure confidentiality.
[2] This ALPE exceeds the 95th percentile and was topcoded to 2.5.

Table A3. Block Counts of American Indians for a Sample Tract

	Tract	Block	****AI Block counts*****			
Block	Blks/tract	ALPE	ALPE	AREX	Census	Difference
1234501.47	34	2.5	.	0	0	0
1234502.47	34	2.5	-1	0	1	-1
1234503.47	34	2.5	.	0	0	0
1234504.47	34	2.5	.	0	0	0
1234505.47	34	2.5	.	0	0	0
1234506.47	34	2.5	.	0	0	0
1234507.47	34	2.5	-1	0	1	-1
1234508.47	34	2.5	.	0	0	0
1234509.47	34	2.5	.	0	0	0
1234510.47	34	2.5	.	.	0	.
1234511.47	34	2.5	.	0	0	0
1234512.47	34	2.5	.	0	0	0
1234513.47	34	2.5	.	1	0	1
1234514.47	34	2.5	.	0	0	0
1234515.47	34	2.5	.	0	0	0
1234516.47	34	2.5	.	0	0	0
1234517.47	34	2.5	.	0	0	0
1234518.47	34	2.5	.	0	0	0
1234519.47	34	2.5	.	0	0	0
1234520.47	34	2.5	.	0	0	0
1234521.47	34	2.5	.	0	0	0
1234522.47	34	2.5	.	0	0	0
1234523.47	34	2.5	.	0	0	0
1234524.47	34	2.5	.	0	0	0
1234525.47	34	2.5	.	0	0	0
1234526.47	34	2.5	.	0	0	0
1234527.47	34	2.5	.	0	0	0
1234528.47	34	2.5	.	0	0	0
1234529.47	34	2.5	.	0	0	0
1234530.47	34	2.5	.	0	0	0
1234531.47	34	2.5	.	4	0	4
1234532.47	34	2.5	0.000	1	1	0
1234533.47	34	2.5	.	5	0	5
1234534.47	34	2.5	.	0	0	0
Tract Total				**11**	**3**	**8**

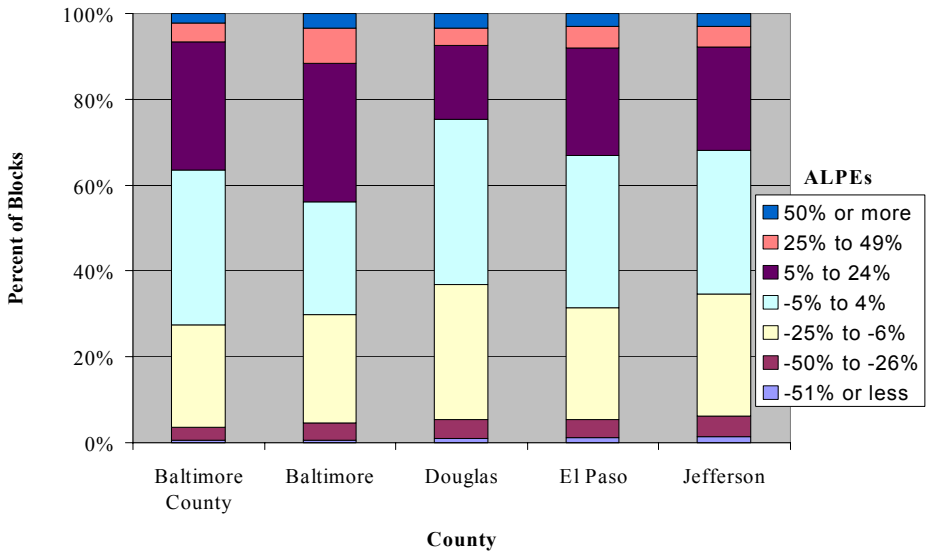
APPENDIX 4: BLOCK-LEVEL ANALYSES

Block-level demographic ALPEs

Summary of Results: The block-level ALPE results provided the least accurate measure of total population (26 to 38 percent of blocks met the five percent criterion and about 85 percent met the 25 percent criterion), compared to tract and county results. But block results were better than tract ALPEs for sex and selected age groups (0-4, 20-24, 65+, older age groups). Race groups with larger populations provided better estimates of Census counts at the five percent criterion, but all block-level ALPEs were worse using the 25 percent criterion. The block-level results exclude zero blocks and mean county ALPEs are affected by smaller denominators, an especially important issue for small population groups that reside in few blocks.

(Figure repeated from section 4.4)

Figure 4.4.1: Distribution of Blocks with Under- and Overcounts of Total Population



The block ALPE results describe the accuracy of counts at the smallest geographic level and relative to counties and tracts. The main problem with this type of comparison is the ALPE denominator potentially inflates block-level ALPEs for small population subgroups and especially minorities. This inflation is likely to be greater than found in the tract-county comparisons. A second issue affecting comparisons is the exclusion of blocks where census did not identify persons with a particular attribute (zero blocks). Tract and block ALPEs include blocks with zero counts because these blocks were collapsed into larger geographies. However, the block-level ALPEs use the reduced sample of blocks and the results may be quite different when comparing the ALPEs at various geographies.

TOTAL POPULATION

- AREX was more accurate in estimating tracts than blocks in all counties; from 26 to 38 percent of blocks were within the five percent criterion, and about 85 percent were within the 25 percent criterion in the five counties; Douglas County had the best results at the five percent criterion and Baltimore County was best at the 25 percent criterion.
- In the MD counties, slightly more blocks had moderate or large overcounts (ALPEs exceeding five percent, compared to the CO counties where more blocks had moderate undercounts (minus five percent to –24 percent; distributions not shown).

The AREX counts were less accurate at the block-level. Total population proportions are likely to be less accurate at smaller areas due to incorrect assignment of households at tracts and blocks that average out for county-level counts. This is demonstrated by the greater number of moderate and large ALPEs and indicates how smaller denominators and AREX processing flaws influenced the results. Though zero blocks were excluded and fewer blocks met the five percent criterion, a surprisingly large proportion of blocks met the 25 percent criterion in all five counties.

SEX

Figure A4.1a: Proportion of Blocks With Sex ALPEs Below 5% and 25% - Baltimore County

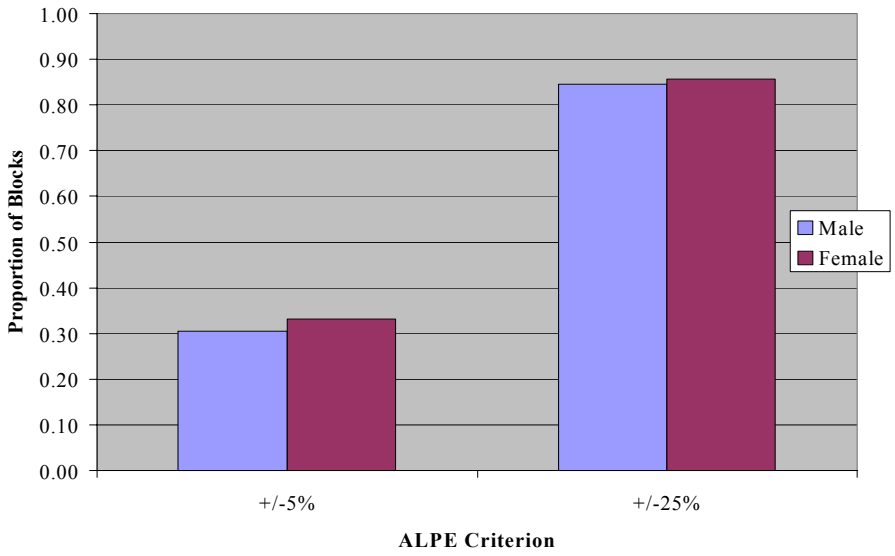


Figure A4.1b: Proportion of Blocks With Sex ALPEs Below 5% and 25% - Baltimore City

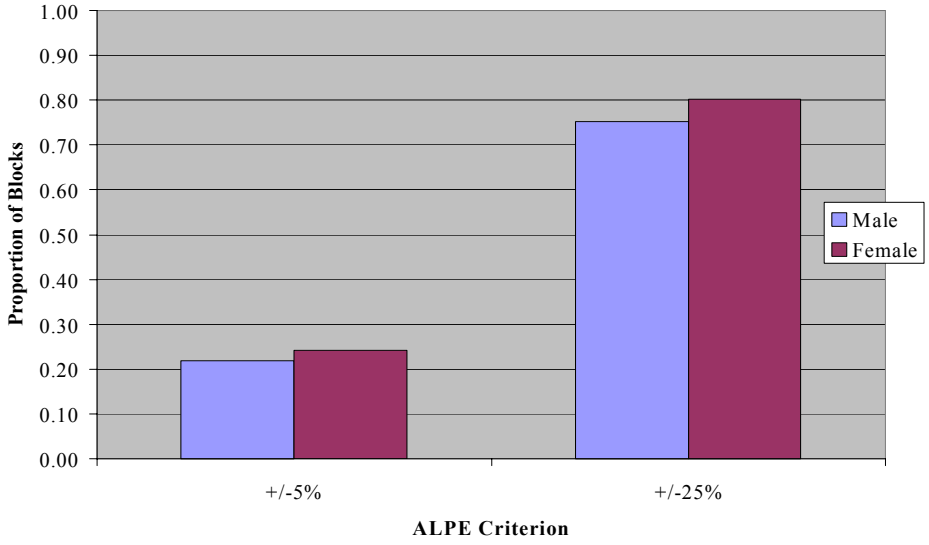


Figure A4.1c: Proportion of Blocks With Sex ALPEs Below 5% and 25% - Douglas County

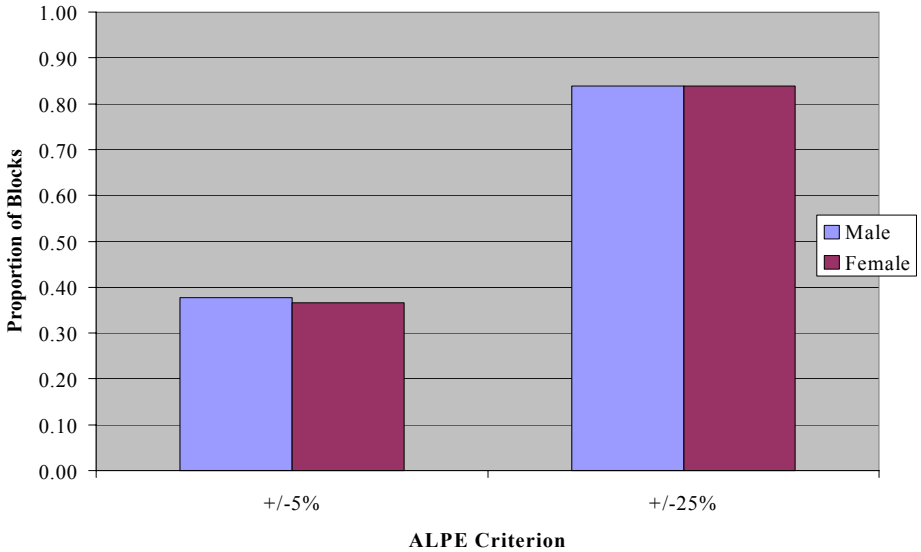


Figure A4.1d: Proportion of Blocks With Sex ALPEs Below 5% and 25% - El Paso County

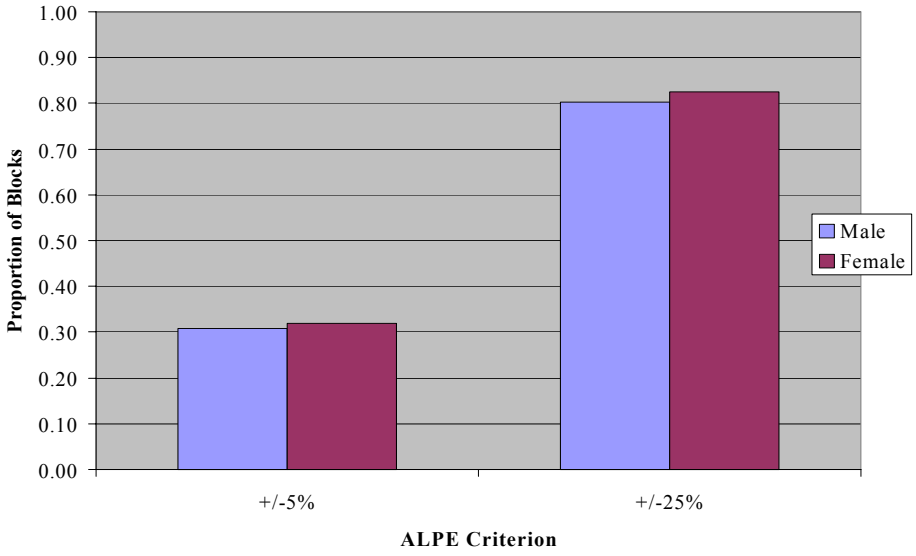
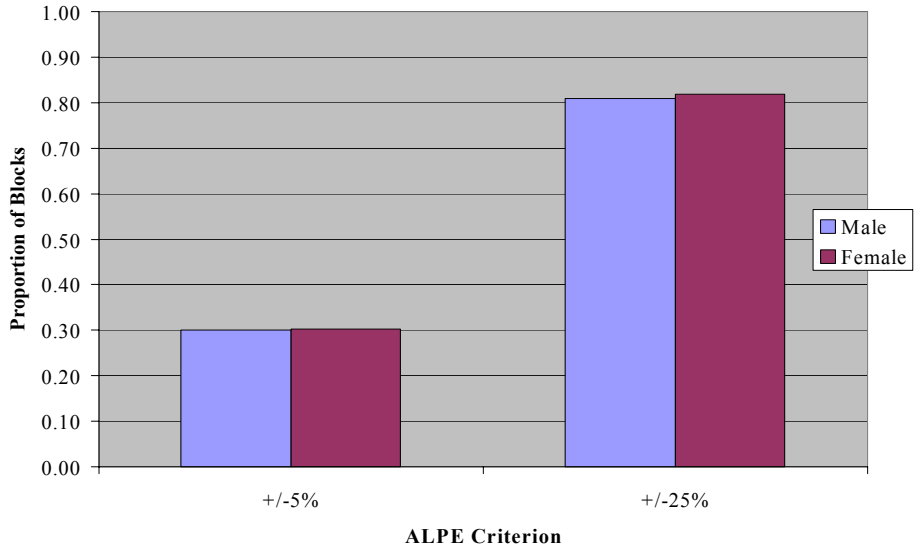


Figure A4.1e: Proportion of Blocks With Sex ALPEs Below 5% and 25% - Jefferson County



- The accuracy of AREX sex results at the five percent criterion was better for blocks than tracts.
- From 39 to 55 percent of male and female ALPEs were within the five percent criterion in the five counties; from 91 to 94 percent of blocks were within the 25 percent criterion.

Male and female undercounts were similar at all geographic levels and reflected the total population results. This similarity suggests that AREX processing was neutral towards whether individuals were male or female. However, males and females have different demographic rates (migration and mortality) at different points in the life-cycle, which may account for the small differences in the male and female AREX results.

AGE

Figure A4.2a: Proportion of Blocks with Age ALPEs Below 5% and 25% - Baltimore County

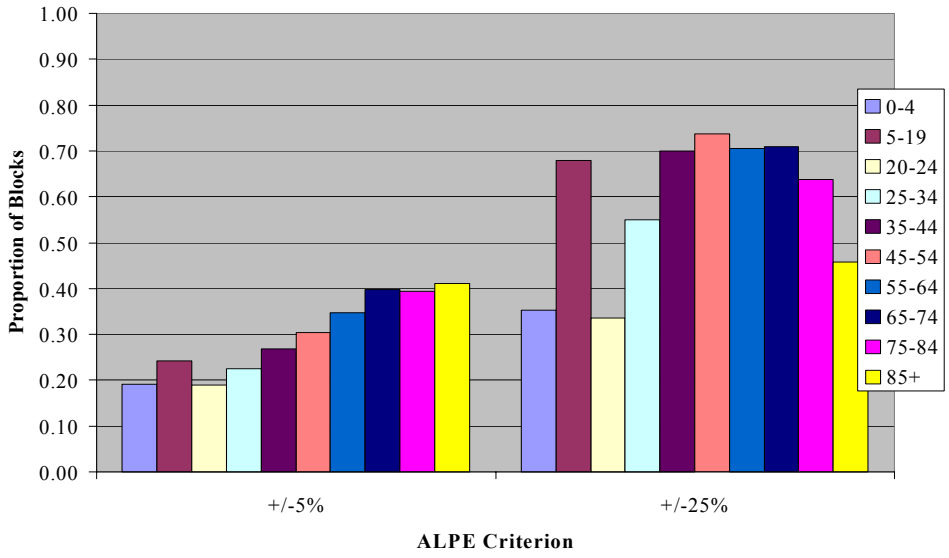


Figure A4.2b: Proportion of Blocks with Age ALPEs Below 5% and 25% - Baltimore City

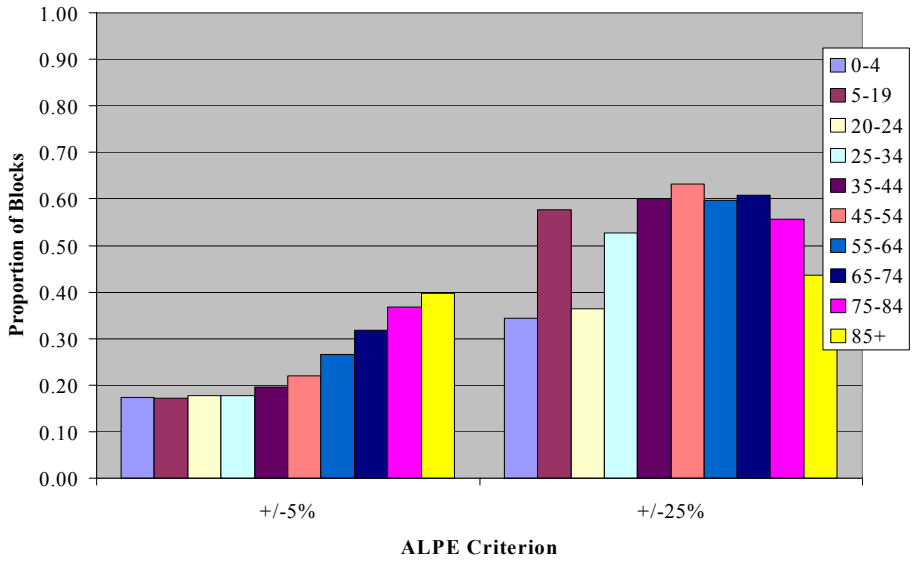


Figure A4.2c: Proportion of Blocks with Age ALPEs Below 5% and 25% - Douglas County

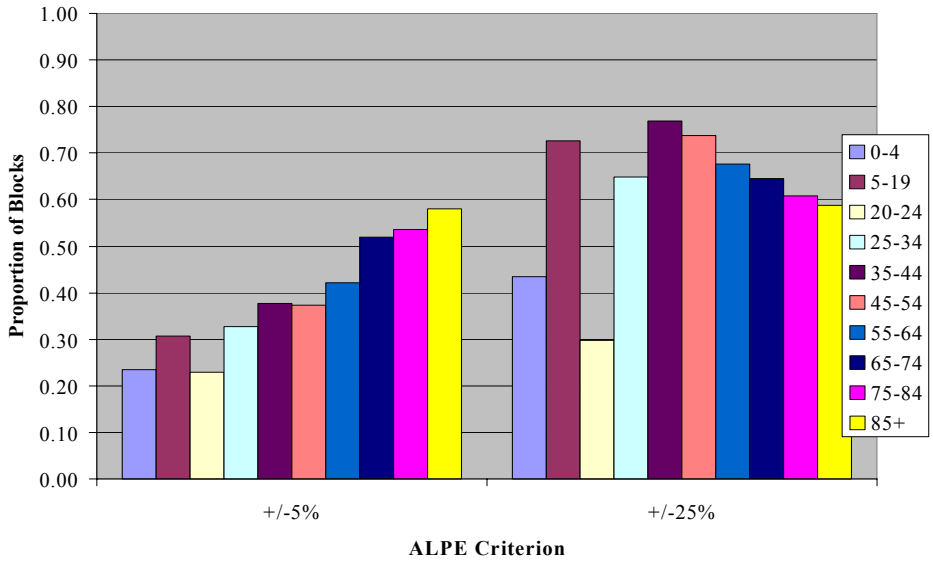


Figure A4.2d: Proportion of Blocks with Age ALPEs Below 5% and 25% - El Paso County

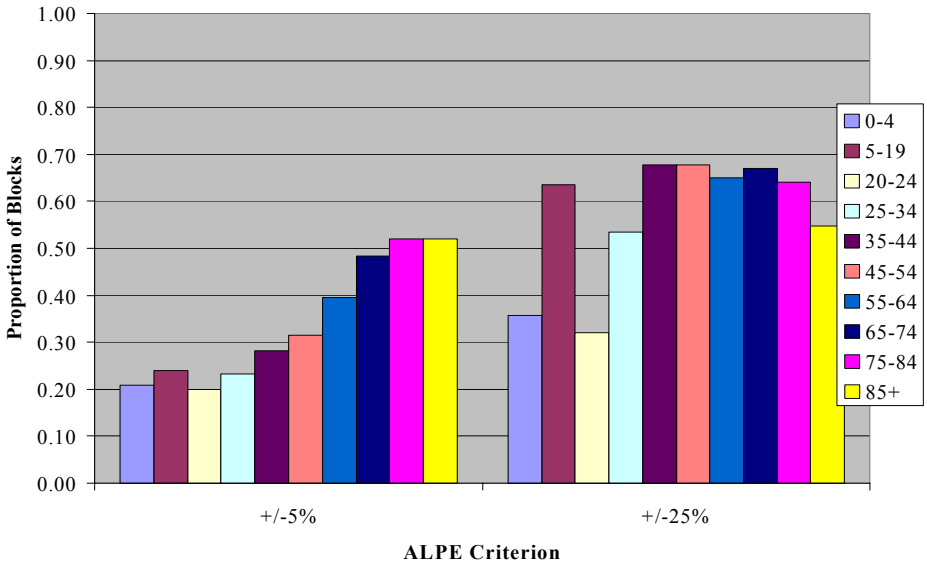
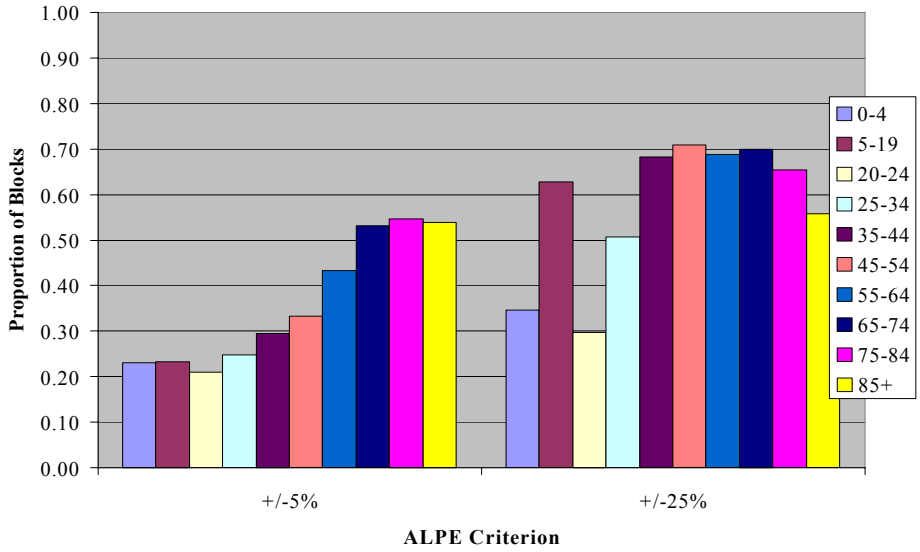


Figure A4.2e: Proportion of Blocks with Age ALPEs Below 5% and 25% - Jefferson County



- Age ALPE results support previous findings from tract and county results: AREX counts were within five percent of Census counts more often for the age 25-74 groups than younger age groups.
- The age ALPE results for age 25-64 age groups were much worse for blocks than tracts in all counties at both five percent and 25 percent criteria; however, block-level results were better for the age 0-4, 20-24, and 65+ age groups at the five percent criterion.
- Old age ALPEs at the five percent criterion were much better for blocks than tracts; though a smaller proportion of blocks had ALPEs of less than five percent, compared to tracts; results for the 75-84 and 85+ age groups were as good or better than for the 65-74 age group.

In general, the block-level results for age were less accurate than the tract-level ALPE results. Besides having smaller denominators for ALPE calculations, blocks with zero population counts are excluded from the analyses. But if AREX performs poorly in some blocks and those blocks are contiguous, it suggests that some block-level ALPE results may be better than corresponding tract ALPEs. That is, errors may be smaller in blocks but cumulated into larger ALPEs within tracts. This may be the case for the 0-4, 20-24, and 65+ age groups because a larger proportion of blocks (compared to tracts) met the five percent criterion.

RACE / ETHNICITY

Figure A4.3a: Proportion of Blocks with Race ALPEs Below 5% and 25% - Baltimore County

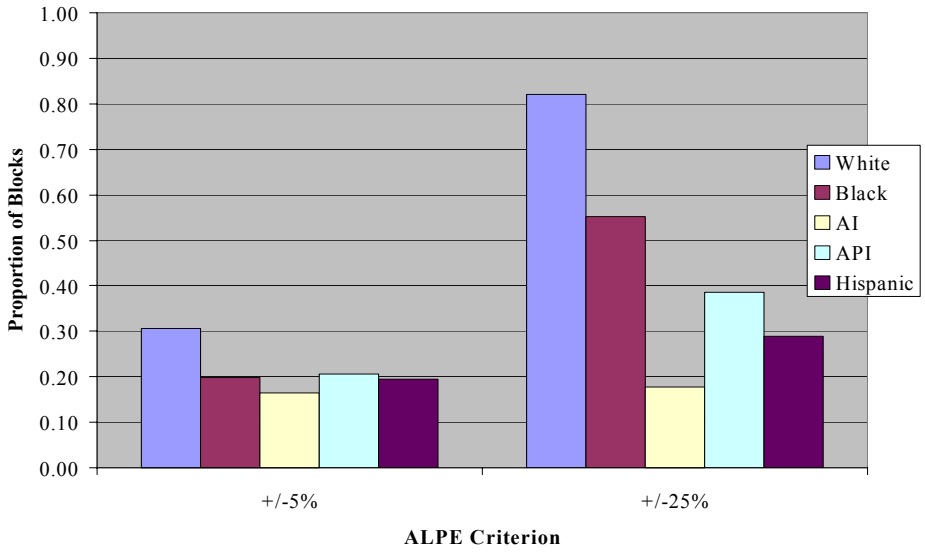


Figure A4.3b: Proportion of Blocks with Race ALPEs Below 5% and 25% - Baltimore City

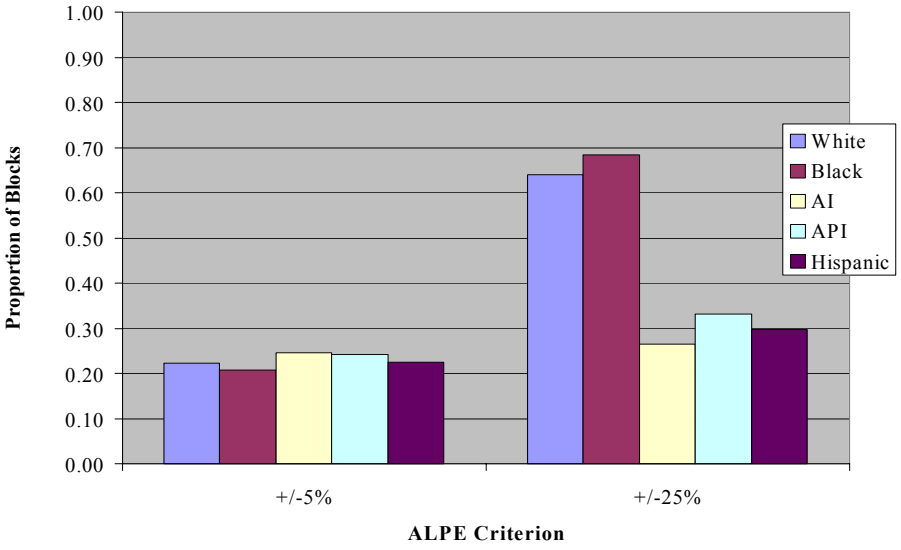


Figure A4.3c: Proportion of Blocks with Race ALPEs Below 5% and 25% - Douglas County

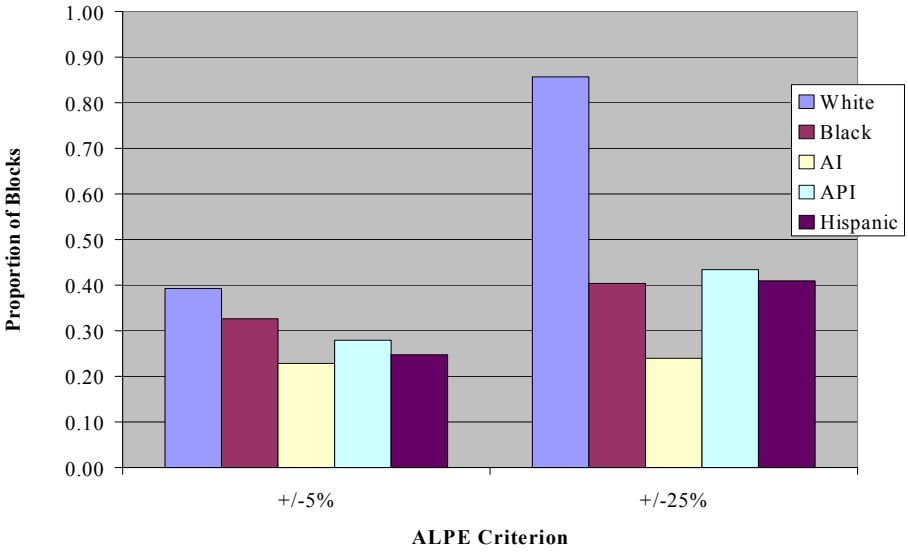


Figure A4.3d: Proportion of Blocks with Race ALPEs Below 5% and 25% - El Paso County

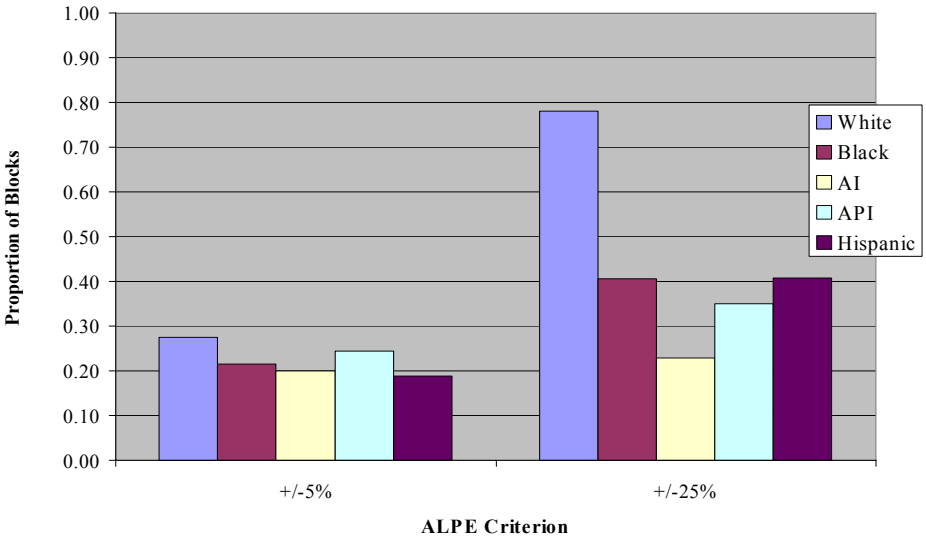
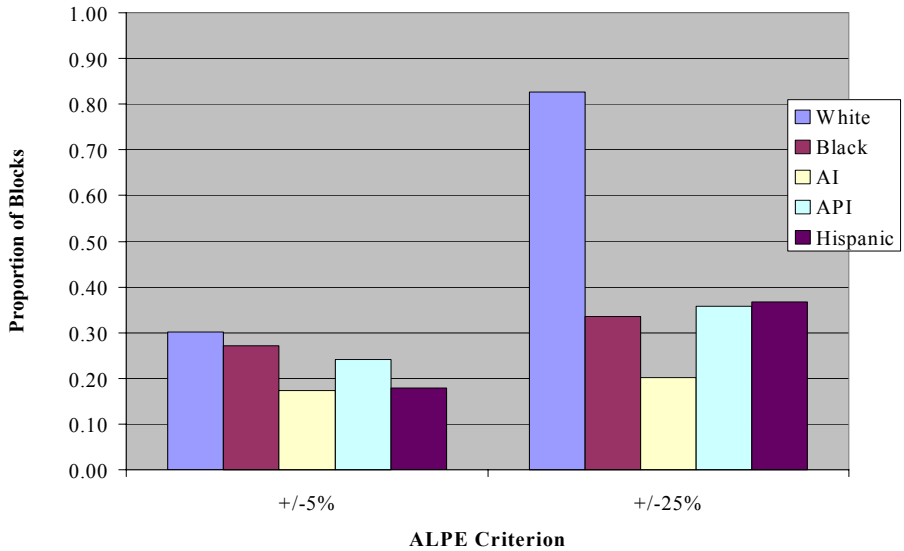


Figure A4.3e: Proportion of Blocks with Race ALPEs Below 5% and 25% - Jefferson County



- In general, ALPE results at the five percent criterion were better for blocks than tracts; but race groups with smaller populations were less accurately counted by AREX.
- All race groups had fewer blocks meeting the 25 percent criterion, compared to tract results.
- In the MD counties, a smaller proportion of blocks were within the five percent criterion for Whites and Blacks, compared to tracts; but a larger proportion of each of the other race groups was within the five percent criterion.

The expected pattern of smaller geography and less accurate AREX counts is supported by the AREX results at the 25 percent criterion. But there is a general tendency for some race groups to be counted more accurately at the block rather than tract-level. This again suggests that cumulative errors may be occurring at tract and county levels, and is especially evident for AIs and APIs.

APPENDIX 5: MULTIVARIATE MODEL PARAMETER ESTIMATES

Table A5.1a: Categorical Logistic Regression Results Predicting Total Block-Level ALPEs-MD (n=13731)[1]

	Large undercount		Moderate undercount		Moderate overcount		Large overcount	
	ß	se	ß	se	ß	se	ß	se
intercept	-1.932 **	0.183	-0.964 **	0.147	-0.923 **	0.153	-2.146 **	0.189
Baltimore City	0.106 **	0.081	-0.165 **	0.064	0.107	0.066	0.722 **	0.086
vacancy rate (> median)	-0.288 **	0.073	-0.060	0.058	0.444 **	0.058	0.811 **	0.073
rental rate (> median)	1.011 **	0.088	0.390 **	0.071	0.050	0.072	0.438 **	0.087
nonrelatives in HH (> median)	0.332 **	0.073	0.174 **	0.058	-0.030	0.057	-0.160 **	0.071
imputed race %-tax (> median)	-0.791 **	0.071	-0.201 **	0.057	0.206 **	0.056	0.173 **	0.070
imputed race %-pcf (> median)	-0.305 **	0.068	-0.010	0.054	0.027	0.053	-0.036	0.067
imputed ethnicity %(> median)	0.851 **	0.124	0.706 **	0.109	0.898 **	0.119	0.880 **	0.133
censpull % (> median)	-0.092	0.077	0.124 **	0.062	0.176 **	0.061	0.080	0.076
multi-race-any mention	-0.268 **	0.073	-0.037	0.057	-0.169 **	0.058	-0.510 **	0.079
some other race-any mention	-0.003	0.091	0.001	0.071	-0.266 **	0.077	-0.215 **	0.113
population density-Q1	0.099	0.108	-0.652 **	0.089	-0.273 **	0.083	0.526 **	0.101
population density-Q2	0.088	0.098	-0.244 **	0.075	-0.175 **	0.074	0.255 **	0.094
population density-Q5	0.162 **	0.082	0.117	0.065	-0.012	0.067	-0.414 **	0.085
neighborhood-1	-0.044	0.138	-0.139	0.109	-0.108	0.105	0.117	0.124
neighborhood-3	0.446 **	0.123	-0.002	0.094	-0.261 **	0.095	-0.033	0.125
neighborhood-4	0.421 **	0.117	-0.020	0.088	-0.130	0.085	-0.039	0.111
White %-Q1	0.243	0.212	-0.714 **	0.199	-0.294	0.207	0.920 **	0.225
White %-Q2	0.096	0.192	-0.619 **	0.181	-0.343	0.194	0.658 **	0.215
White %-Q5	-0.217 **	0.094	-0.175 **	0.074	0.036	0.072	0.422 **	0.094
Blacks-any	-0.358	0.200	0.422 **	0.186	0.633 **	0.198	0.031	0.217
Hispanics-any	-0.316 **	0.080	-0.002	0.061	-0.108	0.061	-0.599 **	0.085
age < 5 (> median)	0.466 **	0.068	0.342 **	0.054	-0.312 **	0.055	-0.596 **	0.071
age 5-19 (> median)	0.149	0.078	0.211 **	0.062	-0.230 **	0.062	-0.564 **	0.078
age 20-24 (> median)	0.398 **	0.070	0.194 **	0.057	0.089	0.057	-0.036	0.071
age 25-44 (> median)	-0.042	0.074	0.173 **	0.059	-0.064	0.059	-0.076	0.073
age 65+ (> median)	-0.265 **	0.077	-0.012	0.062	0.071	0.062	-0.071	0.076

** $p < .05$; used to distinguish important predictors for these non-sample data.

[1]Reference range is best quartile estimate, where AREX is -2.3% to +5.5% of Census total population

Large undercount < -14.4%; Moderate undercount=-14.4% to -2.3%; Moderate overcount=5.5% to 19.8%; Large overcount > 19.8%

Table A5.1b: Categorical Logistic Regression Results Predicting Total Block-Level ALPEs-CO (n=16948)[1]

	Large undercount		Moderate undercount		Moderate overcount		Large overcount	
	β	se	β	se	β	se	β	se
intercept	-2.346 **	0.162	-1.707 **	0.139	-1.363 **	0.149	-2.352 **	0.168
Douglas County	-0.126	0.091	0.180 **	0.071	-0.002	0.075	-0.105	0.093
El Paso County	-0.191 **	0.066	-0.253 **	0.055	0.063	0.054	0.043	0.065
vacancy rate (> median)	-0.057	0.088	0.133	0.074	0.550 **	0.074	0.740 **	0.084
rental rate (> median)	0.527 **	0.065	0.238 **	0.054	0.011	0.053	0.193 **	0.065
nonrelatives in HH (> median)	0.456 **	0.064	0.305 **	0.052	0.159 **	0.051	-0.074	0.064
imputed race %-tax (> median)	-0.651 **	0.068	-0.007	0.055	0.355 **	0.054	0.331 **	0.067
imputed race %-pcf (> median)	-0.065	0.064	0.178 **	0.051	0.248 **	0.051	0.521 **	0.064
imputed ethnicity %(> median)	0.936 **	0.099	1.260 **	0.095	1.559 **	0.113	1.393 **	0.113
censpull % (> median)	0.133	0.070	0.342 **	0.058	0.028	0.059	-0.020	0.071
multi-race-any mention	-0.530 **	0.074	-0.219 **	0.056	-0.315 **	0.056	-0.772 **	0.077
some other race-any mention	-0.214 **	0.079	-0.119 **	0.059	-0.182 **	0.059	-0.222 **	0.084
population density-Q1	0.134	0.085	-0.503 **	0.073	-0.318 **	0.070	0.693 **	0.084
population density-Q2	0.172 **	0.077	-0.107	0.061	0.029	0.059	0.468 **	0.078
population density-Q5	0.496 **	0.110	0.299 **	0.089	-0.349 **	0.106	-0.024	0.148
neighborhood-1	0.106	0.111	0.125	0.086	-0.104	0.083	-0.138	0.110
neighborhood-2	0.653 **	0.104	0.170 **	0.086	-0.115	0.086	0.130	0.107
neighborhood-4	0.278 **	0.111	0.030	0.096	-0.302 **	0.092	-0.011	0.102
White %-Q1	0.379	0.332	-2.721 **	1.017	n/a	n/a	0.628 **	0.298
White %-Q2	0.651 **	0.084	-0.026	0.067	-0.095	0.069	0.844 **	0.089
White %-Q5	0.132	0.087	-0.445 **	0.074	-0.445 **	0.072	0.214 **	0.086
Blacks-any	-0.355 **	0.102	-0.130	0.079	0.056	0.080	-0.093	0.105
Hispanics-any	-0.270 **	0.077	0.110	0.065	0.096	0.064	-0.448 **	0.077
age < 5 (> median)	0.537 **	0.067	0.326 **	0.054	-0.291 **	0.055	-0.603 **	0.074
age 5-19 (> median)	0.487 **	0.065	0.271 **	0.055	-0.033	0.055	-0.569 **	0.069
age 20-24 (> median)	0.334 **	0.066	0.149 **	0.055	0.139 **	0.055	0.151 **	0.068
age 25-44 (> median)	-0.145 **	0.065	-0.016	0.054	-0.263 **	0.054	-0.102	0.067
age 65+ (> median)	0.148 **	0.072	0.174 **	0.061	0.183 **	0.059	0.244 **	0.070

** $p < .05$; used to distinguish important predictors for these non-sample data.

[1]Reference range is best quartile estimate, where AREX is -4.2% to +2.0% of Census total population

Large undercount < -16.7%; Moderate undercount= -16.7% to -4.2%; Moderate overcount=2.0% to 16.2%; Large overcount > 16.2%

Table A5.2a: Piecewise Logistic Regression Results Predicting Total Block-Level ALPEs-MD (n=13731)[1]

		Mobility Variables				Imputation/Process variables			Other Race	Population Density Quintile			Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-3
1	-0.289	0.014	-0.016		0.043	0.032		0.026			-0.063	-0.018		-0.027	-0.027
2	-0.068	0.005	-0.004	-0.003	0.004								-0.003		
3	0.110	0.004											-0.005		
4	0.278	0.006			0.005	-0.005			-0.007		0.020	0.012	-0.012		
5	0.010					0.003	0.007	0.003			-0.003			-0.004	-0.005

Table A5.2a (cont'd)

	White Quintiles				Race		County	Age Groups				
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Baltimore	< 5	5-19	20-24	25-44	65+
1	-0.029	-0.040	-0.035		0.036		-0.016					0.032
2						0.003				-0.003		0.003
3		0.013				-0.005	0.006					
4		0.021					0.009	-0.010	-0.006			
5	-0.003	-0.006		-0.003	0.007		-0.002					

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

[1]See notes at end of Table A5.2f for additional information.

Table A5.2b: Piecewise Logistic Regression Results Predicting Total Block-Level ALPEs-CO (n=16948)[1]

		Mobility Variables			Imputation/Process variables			Other Race		Population Density Quintile			Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-2
1	-0.301			0.015	0.056	0.032		0.019			-0.053	-0.030	-0.020		-0.034
2	-0.092		-0.003		0.004								-0.006		-0.005
3	0.089								-0.004	-0.003	0.008			-0.005	
4	0.258			-0.007			-0.013		-0.007		0.022				0.011
5				-0.002			-0.006	-0.002			0.002				

Table A5.2b (cont'd)

	White Quintiles				Race		County	Age Groups					
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Douglas	El Paso	< 5	5-19	20-24	25-44	65+
1	-0.026	-0.100	-0.032		0.024	0.015			-0.014	-0.016		0.013	
2	-0.005		-0.006						-0.003				
3			0.008			-0.005			-0.004	-0.005			
4	0.014	0.052	0.026	0.020		-0.012			-0.007	-0.012	-0.009		
5				0.003		-0.003			-0.002	-0.002			

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

[1]See notes at end of Table A5.2f for additional information.

Table A5.2c: Piecewise Logistic Regression Results Predicting Age 0-4 Block-Level ALPEs-CO (n=12603 for all 5 models) [1]

		Mobility Variables				Imputation/Process variables			Other Race	Population Density Quintile			Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-2
1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
2	-0.543		-0.011		0.026		-0.034	0.010							
3	-0.319				0.013		-0.017	0.013							
4	0.754								-0.163						
5	0.037				-0.011		-0.031	-0.013	-0.022	-0.013	0.008	0.006			

Table A5.2c (cont'd)

	White Quintiles				Race		County		Age Groups				
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Douglas	El Paso	< 5	5-19	20-24	25-44	65+
1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
2				0.013		-0.017							0.014
3				0.010					-0.014	-0.006			
4			0.112		-0.131				-0.430				0.172
5			0.023	0.006	-0.015	-0.018	-0.012		-0.039				0.008

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

[1]See notes at end of Table A5.2f for additional information.

Table A5.2d: Piecewise Logistic Regression Results Predicting Age 65+ Block-Level ALPEs-MD (n=12688 for all 5 models) [1]

		Mobility Variables			Imputation/Process variables				Other Race		Population Density Quintile			Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-3
1	-0.534	0.039			0.029	0.067	0.063	0.048	0.044		-0.155	-0.076		-0.054	-0.091
2	0.098		-0.007						-0.006				-0.005		
3	0.275	0.008													
4	0.741										0.289	0.190			0.193
5							-0.008		-0.005	-0.007	0.010		-0.007		0.007

Table A5.2d (cont'd)

	White Quintiles				Race		County	Age Groups				
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Baltimore	< 5	5-19	20-24	25-44	65+
1	-0.081	-0.164	-0.137		0.110	0.056			0.031			0.107
2						-0.007	0.008					-0.012
3						-0.007		0.007				-0.020
4		0.271										-0.281
5	0.006			0.004		-0.007					0.004	-0.015

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

[1]See notes at end of Table A5.2f for additional information.

Table A5.2e: Piecewise Logistic Regression Results Predicting Block-Level Black ALPEs-MD (n=11238 for all 5 models) [1]

		Mobility Variables				Imputation/Process variables			Other Race		Population Density Quintile			Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-3
1	-0.778					0.048		0.045			-0.069				
2	-0.254		-0.018		0.012				0.021	0.012	-0.020				
3	0.173							-0.009			0.017		-0.013		
4	0.952			0.330	-0.303			-0.323			0.910				
5															

Table A5.2e (cont'd)

	White Quintiles				Race		County	Age Groups				
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Baltimore	< 5	5-19	20-24	25-44	65+
1	-0.046			-0.140		0.042		0.029		0.031		-0.031
2		0.047	0.038	-0.123	0.027	0.011		-0.008				
3					-0.030	-0.011	0.017	-0.007				
4												
5												

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

[1]See notes at end of Table A5.2f for additional information.

Table A5.2f: Piecewise Logistic Regression Results Predicting Block-Level Hispanic ALPEs-CO (n=11289 for all 5 models) [1]

		Mobility Variables				Imputation/Process variables			Other Race	Population Density Quintile				Neighborhoods	
Q-grp	Intercept	vac50	rent50	nrel50	imp-tax	imp-pcf	imp-hisp	censpull	mrace	sor	Q1	Q2	Q5	neigh-1	neigh-2
1	-0.989		0.025	0.013	0.013	0.047			0.030	0.041	-0.026	-0.020	-0.028		
2	-0.472				0.016	0.054		0.014	0.022	0.012	-0.025	-0.016	-0.023	0.019	
3	0.434								-0.047	-0.071			-0.039		
4	1.494				0.336	0.210				-0.303					
5	0.017		-0.009		-0.012	-0.021	-0.020		-0.019	-0.029	0.009				

Table A5.2f (cont'd)

		White Quintiles			Race		County		Age Groups				
Q-grp	neigh-4	Q1	Q2	Q5	Black	Hispanic	Douglas	El Paso	< 5	5-19	20-24	25-44	65+
1					0.026					0.017			
2			-0.017		0.022			-0.014					
3													
4				-0.388									
5							0.009	0.006			-0.006		

note: $p < .05$ for all parameter estimates shown; non-significant estimates have been omitted to simplify display; used to distinguish important predictors for these non-sample data.

Q-grps relative to reference group: 1=large undercount, 2=moderate undercount, 3=moderate overcount, 4=large overcount

vac50=vacant proportion of block housing units: binary indicator of top 50%
rent50=rental proportion of block housing units: binary indicator of top 50%
nrel50=non-relative household member proportion of block households: binary indicator of top 50%
imp-tax=proportion of cases imputed using tax method: binary indicator of top 50%
imp-pcf=proportion of cases imputed using pcf method: binary indicator of top 50%
imp-hisp=proportion of cases imputed for hispanic origin: binary indicator of top 50%
censpull=proportion of census pull cases: binary indicator of top 50%
mrace=presence of multi-race reports

sor=presence of some other race reports
popd Q1, Q2, Q5=population density quintiles of block: Q1 (low), Q2, Q5 with Q3 and Q4 as the reference group
neigh-1, 2, 3, 4=type of neighborhood from factor analyses
wht Q1, Q2, Q5=White quintile groups from White proportion of block population: Q1 (low), Q2, Q5 with Q3 and Q4 as the reference group
Black=Black proportion of block population: binary indicator of top 50%
hispanic=presence of any Hispanic residents on block
baltimore=indicator for Baltimore City: Baltimore County as reference group; douglas, elpaso=indicators for counties: Jefferson County as reference group
age < 5=age 0-4 proportion of block population: binary indicator of top 50%
age 5-19=age 5-19 proportion of block population: binary indicator of top 50%
age 20-24=age 0-4 proportion of block population: binary indicator of top 50%
age 25-44=age 25-44 proportion of block population: binary indicator of top 50%
age 65+=age 65+ proportion of block population: binary indicator of top 50%

APPENDIX 6: GLOSSARY OF TERMS AND DEFINITIONS

ALPE	Algebraic Percent Error, formed from Census and AREX counts using Census results as the standard.
ABI	American Business Information; ABI is a commercially available list of residential and business addresses covering the entire U.S.
AI	American Indians.
API	Asian and Pacific Islanders.
AREX 2000	Administrative Records Experiment in 2000.
Bottom-up	Bottom-up method of processing AREX counts that includes MAF address verification and variable imputation.
Census-pull	For addresses that failed to match the MAF, the bottom-up process replaced some of these addresses with actual Census 2000 records.
Code-1	Code-1 is a commercially available software product used to standardize and match addresses to other address lists.
FAV estimation	For addresses that failed to match the MAF, the bottom-up process replaced some of these addresses using estimated counts derived from a sample of households that were authenticated by a field address verification (FAV) process.
GIS	Geographic information system.
Hispanic origin	Hispanic origin of any type, based on administrative reports, surname processing, country of origin, and Hispanic origin of householder.
Hot deck assignment	The race imputation process used statistical models to calculate expected race probabilities for each person. The hot deck assignment was based on an algorithm that compared the calculated probability with a randomly drawn number to determine whether a calculated probability was large enough to be assigned to a particular race category.
Index of Dissimilarity	Index of summed differences between AREX and Census counts based on either race/ethnicity or age groups.
MAF or Master Address File	The master list of verified household addresses used to conduct Census-related activities.
Multi-race rate	Derived from Census: based on reported number of race responses.
Neighborhood characteristics	Estimated from factor analyses that distinguish four types of AREX neighborhoods in each AREX state; derived from demographic, housing unit, and population density variables.

Non-relative rate	Derived from Census: proportion of households with non-relative members.
NRFU	Nonresponse follow-up; households that could not be enumerated through usual Census enumeration methods.
Numident	The electronic roster of participants in any of the social programs maintained by the Social Security Administration, compiled from SSN applications, name changes, and corrections.
Overcount	AREX counts that are greater than Census counts, expressed as differences or ALPEs.
PCF probability model	The personal characteristics file (PCF) used a probabilistic race imputation methodology based on logistic regression models and hot deck assignment.
Population density	Population per unit area, expressed as persons per square mile.
PRED	Planning, Research, and Evaluation Division.
Race	AREX race values are based on 'generally accepted' race categories that are derived from complex AREX processing rules; Census race measures use self-reported race from Census forms and exclude persons claiming some other race or multi-race.
Rental rate	Derived from Census: proportion of housing units identified as rental units.
Shannon-Wiener Index of Diversity	Summed index of age or race components using AREX-only measures to distinguish regions with more or less diverse populations.
StARS	Statistical Administrative Records System.
Top-down	Top-Down Administrative Records counts that includes block-coding but no further enhancements.
TIGER	Topologically Integrated Geographic and Cartographic Encoding and Referencing database of all U.S. regions and Puerto Rico.
Undercount	AREX counts that are less than Census counts, expressed as differences or ALPEs.
Vacancy rate	Derived from Census: proportion of housing units identified as vacant.